

# LARGE DEVIATIONS FOR HIERARCHICAL SYSTEMS OF INTERACTING JUMP PROCESSES

BOUALEM DJEHICHE  
AND  
ALEXANDER SCHIED

ABSTRACT. We investigate the large deviations principle from the McKean–Vlasov limit for a collection of jump processes obeying a two–level hierarchy interaction. A large deviation upper bound is derived and it is shown that the associated rate function admits a Lagrangian representation as well as a non-variational one. Moreover, it is proved that the admissible paths for the weak solution of the McKean–Vlasov equation enjoy certain strong differentiability properties.

## 0. Introduction.

In this paper, we are interested in a class of large scale interacting systems which are organized into a *multilevel hierarchy*. Each level of this hierarchy consists of a number of interacting subsystems, at a lower level. This type of interactions arise in many fields such as statistical physics (Dyson’s hierarchical model, the two-level Desai–Zwansig model of ferromagnetism), chemical kinetics (multilevel Schlögl model) and epidemiology (SIR-epidemic models with geographic structure) –to mention only few.

Generally, real large scale hierarchical systems consist of a finite number of individual subsystems at each level of the hierarchy. Among other factors, this structure make them exhibit highly complex behaviour and therefore are difficult to study. Nevertheless, as suggested by Dawson (see [D1], [D2] and the references therein), insight into their qualitative behaviour can be obtained by studying the idealized limit as the number of subsystems per level gets large. The approach to this program is to formulate the relevant stochastic models in the context of *measure-valued* Markov processes. Then, the finite population effect is caught by the study of random perturbations, fluctuations and large deviations about the infinite limit. In this context, Dawson and Gärtner [DG1] established a general multilevel large deviations principle for the empirical measure associated with  $M$  independent copies of a sequence  $(\zeta^N)$  of random variables with values in a topological space.

In this paper, we consider the behaviour of some Markovian systems of jump processes obeying a *two-level* hierarchy interaction, which in the limit, as the number of subsystems grows to infinity, satisfy a McKean–Vlasov type equation. This means that the sequence of two–level empirical measures which represent the states

---

*Key words and phrases.* Two–level hierarchical interaction, McKean–Vlasov limit, large deviations, Orlicz space, measure-valued processes, weak interaction, epidemic process.

*Mathematics Subject Classification:* 60J75, 60F10, 60G57, 60K35.

of the finite systems, have a deterministic limit of a specified form which in a sense represents the law of a typical trajectory in the underlying collection of jump processes that are subject to the hierarchy interaction.

In this study, we derive a large deviation upper bound in the McKean–Vlasov limit and obtain two representations of the associated rate function: A non-variational (integral) representation and a Lagrangian one. The methods of the proof we adopt here are based on the Hamiltonian approach to large deviations, which is natural in the context of sample path LDP's, and are rather different from the ones in Dawson and Gärtner [DG1].

In Section 1, we present the model. The main results are stated in Section 2. The proof of the representation theorems of the rate function is given in Section 3, and the large deviation upper bound result on the canonical path space is derived in section 4.

## 1. The jump process.

**1.1 Model.** Let  $\mathbb{E}$  denote a Polish space with Borel  $\sigma$ -field  $\mathcal{B}_{\mathbb{E}}$ . We consider a system of  $M$  colonies each consisting of  $N$  individuals. Let  $X_t^{ij}$  denote the state of the  $j$ th individual in the  $i$ th colony at time  $t \in I := [0, T]$ , for some fixed  $T > 0$ . Assume that, for each  $(i, j)$ ,  $(X_t^{ij}, t \in I)$  is an  $\mathbb{E}$ -valued jump process with jump rate  $\gamma^{ij}(\cdot)$  and jump kernel  $\pi^{ij}$ . The jump intensity measure  $\gamma^{ij}(\mathbf{x}) dt$  gives the jump rate of the  $j$ th individual in the  $i$ th colony when the system is in the state  $\mathbf{x} \in \mathbb{E}^{NM}$ . The jump size distribution  $\pi^{ij}(x, dy)$ ,  $x \in \mathbb{E}$ , is such that if a jump of the  $j$ th individual in the  $i$ th colony at site  $x \in \mathbb{E}$  occurs, then it is of the form  $x \rightarrow dy \subset \mathbb{E}$  with probability  $\pi^{ij}(x, dy)$ . We are going to assume throughout that  $\pi^{ij}(x, dy) = \pi(x, dy)$ , for a fixed given probability transition kernel  $\pi$  on  $\mathbb{E} \times \mathcal{B}_{\mathbb{E}}$ . Consequently, all interaction is guided into the system through the jump rates  $\gamma^{ij}$ , what means that we are dealing with a jump process analogue of a system of interacting diffusions.

For each colony  $i$ , we can construct the corresponding flow of empirical distributions

$$\mu_t^i = \frac{1}{N} \sum_{j=1}^N \delta_{X_t^{ij}} \quad (t \in I) \quad (1.1)$$

as stochastic process taking values in the space  $\mathcal{M} = \mathcal{M}(\mathbb{E})$  of Borel probability measures on  $\mathbb{E}$ . The *two-level* empirical process then is given by

$$\xi_t^{M,N} = \frac{1}{M} \sum_{i=1}^M \delta_{\mu_t^i} \quad (t \in I). \quad (1.2)$$

It takes values in the space  $\mathcal{M}^{II} = \mathcal{M}(\mathcal{M})$  of probability measures on  $\mathcal{M}$ . Both  $\mathcal{M}$  and  $\mathcal{M}^{II}$  will be endowed with the corresponding weak topologies and Borel  $\sigma$ -fields. Of course, as long as the number  $M$  of colonies is fixed,  $\xi_t$  could be obtained from a standard empirical distribution on the enlarged state space  $\{1, \dots, M\} \times \mathbb{E}$ . However, the two-level model allows the number of colonies to grow simultaneously with the total number of particles, which seems to be natural for many models. See e.g. our example of an SIR epidemic model in Section 2.3. For this reason,

we are interested in an interaction that is guided into the system through the two-level empirical distribution  $\xi$  of the state  $\mathbf{x} = (x^{ij}) \in \mathbb{E}^{NM}$  of the system, and the empirical distribution  $\mu^i$  of the corresponding colony. This means that the jump intensities  $(\gamma^{ij}(\mathbf{x}))$  have to be of the form

$$\gamma^{ij}(\mathbf{x}) = \gamma(x^{ij}, \mu^i, \xi) \quad (1 \leq i \leq M, 1 \leq j \leq N), \quad (1.3)$$

where  $\gamma$  is a fixed function on  $\mathbb{E} \times \mathcal{M} \times \mathcal{M}^{II}$ . For simplicity, we will make the following assumptions on the jump characteristics  $\pi$  and  $\gamma$ :

**Assumption I.** *The kernel  $\pi$  possesses the Feller property.*

**Assumption II.**  *$\gamma$  is a bounded positive and continuous function on  $\mathbb{E} \times \mathcal{M} \times \mathcal{M}^{II}$ .*

In the sequel we will be interested in the asymptotics of the two-level empirical process  $\xi_t = \xi_t^{N,M}$  as  $N, M \rightarrow \infty$ . Due to (1.3),  $\xi_t^{N,M}$  is a measure-valued Markov process. It has an infinitesimal generator  $\mathcal{A}^{N,M}$  acting on bounded measurable functions  $F$  on  $\mathcal{M}^{II}$  as follows:

$$\begin{aligned} \mathcal{A}^{M,N} F(\xi) &= MN \int \bar{m}(\xi, d\mu, dx, dy) \\ &\quad \left[ F\left(\xi + \frac{1}{M} \{\delta_{(\mu + \frac{1}{N}(\delta_y - \delta_x))} - \delta_\mu\}\right) - F(\xi) \right], \end{aligned} \quad (1.4)$$

where we have used the abbreviations

$$\bar{m}(\xi, d\mu, dx, dy) = \xi(d\mu) \mu(dx) m(x, \mu, \xi, dy) \quad (1.5)$$

and

$$m(x, \mu, \xi, dy) = \gamma(\xi, \mu, x) \pi(x, dy). \quad (1.6)$$

Theorems 4.7.3 and 4.4.2 in Ethier and Kurtz [EK] and Assumption II show that the martingale problem associated with  $\mathcal{A}^{N,M}$  is well-posed and that its unique solution is a strong Markov process.

As usual, we will frequently denote the integration of a bounded and measurable function  $f$  on  $\mathbb{E}$  against a measure  $\mu \in \mathcal{M}$  by  $\langle \mu, f \rangle$ . To distinguish it from an integral of a function  $F$  on  $\mathcal{M}^{II}$  with respect to some  $\xi \in \mathcal{M}^{II}$ , we will write  $\ll \xi, F \gg$  in the latter case.

Now choose a totally bounded metric  $d$  generating the topology of  $\mathbb{E}$ . See Parthasarathy [Pa], p. 43, for existence. We define a space  $\mathcal{F}_0$  of finitely based functions on  $\mathcal{M}$  as the set of all functions  $f$  that are of the form

$$f(\mu) = \varphi(\langle \mu, \psi_1 \rangle, \dots, \langle \mu, \psi_k \rangle) \quad (\mu \in \mathcal{M}), \quad (1.7)$$

for some  $k \in \mathbb{N}$ ,  $\varphi \in C_c^1(\mathbb{R}^k)$ , the space of continuously differentiable functions having compact support in  $\mathbb{R}^k$ , and  $\psi_1, \dots, \psi_k \in C_u(\mathbb{E})$ , the space of bounded and uniformly continuous functions on the metric space  $(\mathbb{E}, d)$ . For the ease of notation we will adopt the convention of rewriting (1.7) by

$$f(\mu) = \varphi(\langle \mu, \vec{\psi} \rangle) \quad (\mu \in \mathcal{M}).$$

We will also need a space  $\mathcal{F}$  of time dependent test functions on  $I \times \mathcal{M}$  being of the form

$$f_t(\mu) = f(t, \mu) = \varphi(t, \langle \mu, \vec{\psi}(t) \rangle) \quad (t \in I, \mu \in \mathcal{M})$$

with  $\varphi \in C_b^1(I \times \mathbb{R}^k)$  and a vector  $\vec{\psi} = (\psi_1, \dots, \psi_k)$  such that each component is a function from  $I$  to  $C_u(\mathbb{E})$ , differentiable with respect to the sup-norm  $\|\cdot\|$  on  $C_u(\mathbb{E})$ . Clearly, functions  $f \in \mathcal{F}$  are differentiable with respect to time with derivative

$$\partial_t f(t, \mu) = (\partial_t f_t)(\mu) = (\partial_t \varphi)(t, \langle \mu, \vec{\psi}(t) \rangle) + (\nabla \varphi)(t, \langle \mu, \vec{\psi}(t) \rangle) \cdot \langle \mu, \partial_t \vec{\psi}(t) \rangle.$$

Here  $\nabla$  denotes the usual Euclidean gradient and “ $\cdot$ ” the Euclidean scalar product. In “space direction” we define difference operators  $D^N$  on  $\mathcal{F}_0$  by

$$D^N f(\mu, x, y) = N \left( f\left(\mu + \frac{1}{N}(\delta_y - \delta_x)\right) - f(\mu) \right). \quad (1.8)$$

Then, as  $N \rightarrow \infty$ ,

$$D^N f(\mu, x, y) \longrightarrow Df(\mu, x, y) := (\nabla \varphi)(\langle \mu, \vec{\psi}(t) \rangle) \cdot (\vec{\psi}(y) - \vec{\psi}(x)). \quad (1.9)$$

To any kernel  $k(x, \mu, \xi, dy)$  on  $\mathbb{E} \times \mathcal{M} \times \mathcal{M}^{II} \times \mathcal{B}_{\mathbb{E}}$  and  $\xi \in \mathcal{M}^{II}$  there correspond linear operators  $\mathcal{G}_k^N(\xi)$  ( $N \in \mathbb{N}$ ) and  $\mathcal{G}_k(\xi)$  on  $\mathcal{F}_0$  given by

$$\mathcal{G}_k^N(\xi) f(\mu) = \int \int D^N f(\mu, x, y) k(x, \mu, \xi, dy) \mu(dx)$$

and

$$\mathcal{G}_k(\xi) f(\mu) = \int \int Df(\mu, x, y) k(x, \mu, \xi, dy) \mu(dx)$$

With the above notations, for every  $f \in \mathcal{F}$ ,

$$\mathcal{A}^{N,M} \ll \cdot, f \gg = \ll \cdot, \mathcal{G}_m^N(\cdot) f \gg, \quad (1.10)$$

where  $m$  is the kernel defined in (1.6).

## 2. Main Results.

**2.1 A large deviation upper bound.** Suppose that, for all  $N$  and  $M$ , we are given some deterministic  $\nu^{N,M} \in \mathcal{M}^{II}$  having the form of a two-level empirical distribution as in (1.2). With  $\mathcal{P}^{N,M}$  we then denote the law of the solution of the martingale problem associated with the generator  $\mathcal{A}^{N,M}$  of (1.4) and starting point  $\nu^{N,M}$ . Then  $\mathcal{P}^{N,M}$  is Borel probability measure on the space  $D(I, \mathcal{M}^{II})$  of all càdlàg functions endowed with the Skorohod topology. This topology is independent of an a priori choice of metric on  $\mathcal{M}^{II}$  see Jakubowski [Ja].

The law of large numbers of the system as  $N, M \rightarrow \infty$  has been studied by Dawson [D1]. There the following result has been stated.

**Proposition 1.1.** *Suppose  $\nu^{M,N}$  converges weakly to some  $\nu \in \mathcal{M}^{II}$ . Then  $\mathcal{P}^{M,N}$  converges weakly on  $D(I, \mathcal{M}^{II})$  to  $\delta_\eta$ , where  $\eta$  is the unique deterministic path in  $C(I, \mathcal{M}^{II})$  satisfying  $\eta_0 = \nu$  and solving the weak McKean-Vlasov equation*

$$\begin{aligned} \ll \eta_t, f_t \gg &= \ll \eta_0, f_0 \gg + \int_0^t \ll \eta_s, \partial_s f_s \gg ds \\ &+ \int_0^t \ll \eta_s, \mathcal{G}_m(\eta_s) f_s \gg ds \quad (f \in \mathcal{F}, 0 \leq t \leq T). \end{aligned}$$

Let  $\mathcal{H}^{M,N}$  be the *Hamiltonian* associated to the generator  $\mathcal{A}^{M,N}$  defined, for  $\zeta \in \mathcal{M}^{II}$  and test functions  $f \in \mathcal{F}$ , as

$$\begin{aligned} \mathcal{H}^{M,N}(\zeta, f) &= e^{-\ll \zeta, f \gg} \mathcal{A}^{M,N} e^{\ll \zeta, f \gg} \\ &= MN \int \bar{m}(\zeta, d\mu, dx, dy) \left\{ \exp \left[ \frac{1}{MN} D^N f(\mu, x, y) \right] - 1 \right\}. \end{aligned} \quad (2.1)$$

Define the *scaled Hamiltonian*  $\mathcal{H}$  by

$$\begin{aligned} \mathcal{H}(\zeta, f) &= \lim_{M,N \rightarrow \infty} \frac{1}{MN} \mathcal{H}^{M,N}(\zeta, MNf) \\ &= \int \bar{m}(\zeta, d\mu, dx, dy) \left\{ \exp[Df(\mu, x, y)] - 1 \right\}. \end{aligned}$$

We introduce, for  $0 \leq t \leq T$ , the functionals  $J_t$  acting on  $f \in \mathcal{F}$  and  $\eta \in D(I, \mathcal{M}^{II})$ :

$$J_t(\eta, f) = \ll \eta_t, f_t \gg - \ll \eta_0, f_0 \gg - \int_0^t (\ll \eta_r, \partial_r f_r \gg - \mathcal{H}(\eta_r, f_r)) dr. \quad (2.2)$$

Then, for any  $\nu \in \mathcal{M}^{II}$  and  $\eta \in D(I, \mathcal{M}^{II})$ , let

$$S_\nu(\eta) = \begin{cases} \sup \left\{ J_T(\eta, f) \mid f \in \mathcal{F} \right\} & \text{if } \eta_0 = \nu, \\ \infty & \text{otherwise.} \end{cases}$$

The functional  $S_\nu$  is in fact the rate function governing the large deviation upper bound of our particle system:

**Theorem I.** *Suppose that Assumptions I and II are satisfied and that the starting points  $\nu^{M,N}$  converge weakly to some  $\nu \in \mathcal{M}^{II}$  as  $N, M \rightarrow \infty$ .*

*Then*

$$\limsup_{M,N \rightarrow \infty} \frac{1}{MN} \log \mathcal{P}^{M,N}(F) \leq - \inf_{\eta \in F} S_\nu(\eta),$$

*for each closed subset  $F$  of  $D(I, \mathcal{M}^{II})$ .*

Using a suitable change of measure it is possible to get a partial large deviation lower bound with the same rate function, but restricted to very regular admissible paths. To extend the lower bound to general admissible paths is also possible but requires another type of techniques like variational convergence. This topic is not included in this paper and will appear elsewhere.

Our next goal is to derive additional representation formulae for the rate function  $S_\nu$ .

**2.2 Representations of the rate function and regularity of paths.** We will now define a subset  $H_\nu$  of admissible paths in the space  $C(I, \mathcal{M}^{II})$  of continuous  $\mathcal{M}^{II}$ -valued mappings.

**Definition I.** Fix  $\nu \in \mathcal{M}^{II}$ . Let  $H_\nu$  denote the set of all paths  $\eta \in C(I, \mathcal{M}^{II})$  such that  $\eta_0 = \nu$  and which satisfy a weak McKean-Vlasov equation

$$\begin{aligned} \ll \eta_t, f_t \gg &= \ll \eta_0, f_0 \gg + \int_0^t \ll \eta_s, \partial_s f_s \gg ds \\ &+ \int_0^t \ll \eta_s, \mathcal{G}_{k_s}(\eta_s) f_s \gg ds \quad (f \in \mathcal{F}, 0 \leq t \leq T), \end{aligned} \quad (2.3)$$

for kernels  $k_s(x, \mu, \eta, dy)$  ( $s \in I$ ) which are absolutely continuous with respect to  $m(x, \mu, \eta, dy)$  and who possess a Radon-Nikodym derivative

$$g_s(\xi, \mu, x, y) := \frac{k_s(\xi, \mu, x, dy)}{m(\xi, \mu, x, dy)}$$

which satisfies

$$\int_0^T dt \int d\bar{m}(\eta_t) (g_t \log g_t - g_t + 1) < \infty. \quad (2.4)$$

The next result gives a non-variational representation formula for  $S_\nu$ . It relies essentially on work by Léonard [L1].

**Theorem II.** Fix  $\nu \in \mathcal{M}^{II}$  and  $\eta \in D(I, \mathcal{M}^{II})$ . Then

$$S_\nu(\eta) < \infty \text{ if and only if } \eta \in H_\nu.$$

And in the latter case the following representation holds

$$S_\nu(\eta) = \int_0^T dt \int d\bar{m}(\eta_t) (g_t \log g_t - g_t + 1). \quad (2.5)$$

**Remark.** The fact that  $S_\nu$  is concentrated on continuous paths implies that the large deviation upper bound of Theorem I extends even to subsets  $F$  that are closed with respect to the stronger *uniform topology* on  $D(I, \mathcal{M}^{II})$  — provided that  $F$  is measurable with respect to all  $\mathcal{P}^{N,M}$ . The argument is the same as in Section 18 in Billingsley [Bi].

It turns out that paths  $\eta \in H_\nu$  enjoy certain differentiability properties. To this end, endow the space  $\mathcal{F}_0$  with the norm

$$\|f\|_{\mathcal{F}_0} := \sup_{\mu} |f(\mu)| + \sup_{\mu, x, y} |Df(\mu, x, y)| \quad (f \in \mathcal{F}_0).$$

The set of all finite signed measures on  $\mathcal{M}$  then can be regarded as a subspace of the topological dual  $\mathcal{F}'_0$  of  $\mathcal{F}_0$  by taking the integral of a function in  $\mathcal{F}_0$  against a measure on  $\mathcal{M}$ . We extend the notation  $\ll \vartheta, f \gg$  to denote the duality relation between  $\vartheta \in \mathcal{F}'_0$  and  $f \in \mathcal{F}_0$ . For  $\vartheta \in \mathcal{F}'_0$  and  $\zeta \in \mathcal{M}^{II}$ , we denote with  $\mathcal{L}(\vartheta, \zeta)$  the Legendre transform of  $\mathcal{H}(\zeta, \cdot)$ :

$$\mathcal{L}(\vartheta, \zeta) = \sup_{f \in \mathcal{F}_0} (\ll \vartheta, f \gg - \mathcal{H}(\zeta, f)).$$

**Theorem III.** *If  $\eta \in H_\nu$ , then the mapping  $t \mapsto \eta_t \in \mathcal{F}_0^l$  is strongly differentiable Lebesgue-almost everywhere with derivative  $\dot{\eta}_t$ . Moreover,  $\eta$  can be represented as*

$$\eta_t = \nu + \int_0^t \dot{\eta}_s ds \quad (0 \leq t \leq T),$$

where the integral is to be taken in the sense of Bochner. Moreover, the following Lagrangian representation of the rate function holds

$$S_\nu(\eta) = \int_0^T \mathcal{L}(\dot{\eta}_t, \eta_t) dt. \quad (2.6)$$

**Remark.** The same method we use here to prove Theorem III can be applied to the weak McKean-Vlasov equations relating to the one-level large deviations as considered in Léonard [L2] or Djehiche and Kaj [DK]. There, the operators corresponding to  $\mathcal{G}_k$  are *bounded* linear operators on the space  $C_b(\mathbb{E})$ . This guarantees that the derivatives of the one-level paths exist even as finite signed measures and with respect to the total variation norm. Let us briefly give an example for this loss of regularity in the two-level case. If we assume that there is *no interaction between the colonies*, i.e.  $\gamma$  does not depend on  $\xi$ , Theorem 2.2 of Dawson and Gärtner [DG1] together with the main result in Djehiche and Kaj [DK] yield a large deviation principle with rate function

$$\bar{S}_\nu(\eta) = \inf \left\{ \int R dQ \mid Q \in \mathcal{M}(D(I, \mathcal{M})) \text{ with marginal flow } (\eta_t) \right\}.$$

Here  $R$  denotes the one-level rate function of Djehiche and Kaj [DK]. Consequently, if  $\mu(\cdot) \in D(I, \mathcal{M})$  is such that  $R(\mu) < \infty$ , the two-level path given by  $\eta(t) = \delta_{\mu(t)}$ , ( $t \in I$ ) satisfies  $S_{\eta(0)}(\eta) \leq \bar{S}_{\eta(0)}(\eta) < \infty$ . This simple example shows that one cannot expect much further regularity of the paths in  $H_\nu$  than stated in Theorem II.

**2.3. Example: Epidemic SIR-model.** A typical situation where two-level hierarchy systems are relevant is when modeling the spread of a disease, taking into account the geographic structure of the population. Consider a *closed* population of  $MN$  individuals organized into  $M$  houses, villages or colonies each containing  $N$  individuals. Let the state space  $\mathbb{E}$  of individuals at lower level, informally described as  $\mathbb{E} = [susc, inf, rem] \times position$ , consist of pointers to one of three possible types,  $\mathcal{S} = susceptible$ ,  $\mathcal{I} = infective$  and  $\mathcal{R} = removed$ , and in addition of a position variable  $r_{ij} \in \mathbb{R}$  if individual  $j$  of the  $i$ th colony is of type  $\mathcal{S}$  or  $\mathcal{I}$  and a cemetery position  $\dagger$  if of type  $\mathcal{R}$ . Then, the total number of individuals in the  $i$ th group can be divided into subgroups according to their type, and represented by means of the associated two-level empirical measure:

$$\zeta^{M,N} = \frac{1}{M} \sum_{i=1}^M \delta_{\mu^i},$$

where

$$\mu^i = \frac{1}{N} \sum_{j=1}^N \delta_{X^{ij}} = \frac{1}{N} \left( \sum_{j=1}^{|\mathcal{S}_i|} \delta_{r_{ij}}^{(\mathcal{S})} + \sum_{j=1}^{|\mathcal{I}_i|} \delta_{r_{ij}}^{(\mathcal{I})} + \sum_{j=1}^{|\mathcal{R}_i|} \delta_{r_{ij}}^{(\dagger)} \right).$$

Here,  $|\mathcal{S}_i|$  denotes the number of susceptible individuals in the  $i$ th colony etc.. Since the population is closed, the epidemic process can be determined by the respective processes associated to the susceptible and infective populations. This is denoted by  $\xi^{M,N} = (\mathcal{S}^{M,N}, \mathcal{I}^{M,N})$  and regarded as an element of  $D(I, \mathcal{M}^{II} \times \mathcal{M}^{II})$ .

Suppose  $\lambda(r, r')$  is a nonnegative smooth function on  $\mathbb{R}^2$  and define jump rates by

$$\gamma^{ij}(\text{susc}, \text{inf}; r) = \gamma^{ij}(\text{susc}; r) = \frac{1}{M} \sum_{l=1}^M \frac{1}{N} \sum_{k=1}^{|\mathcal{I}_l|} \lambda(r_{ij}, r_{kl}).$$

for a jump  $\mathcal{S} \rightarrow \mathcal{I}$  of particle  $j$  of colony  $i$  at  $r_{ij}$  and

$$\gamma^{ij}(\text{inf}, \text{rem}; r) = \gamma^{ij}(\text{inf}; r) = \rho$$

for a jump  $\mathcal{I} \rightarrow \mathcal{R}$ . Then the infinitesimal generator of the epidemic process  $\xi^{M,N}$  reads

$$\begin{aligned} \mathcal{A}^{M,N} F(\xi) &= MN \int \mathcal{S}(d\mu_1) \mathcal{I}(d\mu_2) \int \mu_1(dx) \mu_2(dy) \lambda(x, y) \\ &\quad [F(\mathcal{S} + \frac{1}{M} \{\delta_{(\mu_1 - \frac{1}{N} \delta_x)} - \delta_{\mu_1}\}, \mathcal{I} + \frac{1}{M} \{\delta_{(\mu_2 + \frac{1}{N} \delta_y)} - \delta_{\mu_2}\}) - F(\mathcal{S}, \mathcal{I})] \\ &\quad + \rho MN \int \mathcal{I}(d\mu_2) \int \mu_2(dy) [F(\mathcal{S}, \mathcal{I} + \frac{1}{M} \{\delta_{(\mu_2 - \frac{1}{N} \delta_y)} - \delta_{\mu_2}\}) - F(\mathcal{S}, \mathcal{I})], \end{aligned}$$

and, with

$$\ll \eta, f \gg = \ll \mathcal{S}, f_1 \gg + \ll \mathcal{I}, f_2 \gg$$

for  $f = (f_1, f_2) \in \mathcal{F} \times \mathcal{F}$ , the corresponding scaled Hamiltonian is,

$$\begin{aligned} \mathcal{H}(\xi, f) &= \int \mathcal{S}(d\mu_1) \mathcal{I}(d\mu_2) \int \mu_1(dx) \mu_2(dy) \lambda(x, y) \\ &\quad [\exp(\nabla \varphi_2(\mu_2) \cdot \vec{\psi}_2(y) - \nabla \varphi_1(\mu_1) \cdot \vec{\psi}_1(x)) - 1] \\ &\quad + \rho \int \mathcal{I}(d\mu_2) \int \mu_2(dy) [\exp(-\nabla \varphi_2(\mu_2) \cdot \vec{\psi}_2(y)) - 1]. \end{aligned}$$

Along the extremals of the corresponding variational problem for the Lagrangian  $\mathcal{L}$ , there are functions  $g^1$  and  $g^2$  such that the corresponding rate function can be written as

$$S(\eta) = \int_0^T \int \mathcal{S}_r(d\mu_1) \mathcal{I}_r(d\mu_2) \left( \langle \mu_1 \otimes \mu_2, \lambda(g_r^1 \log g_r^1 - g_r^1 + 1) \rangle + \rho \langle \mu_2, (g_r^2 \log g_r^2 - g_r^2 + 1) \rangle \right) dr,$$

which agrees with (2.5).

### 3. Proof of Theorems II and III.

**3.1 Proof of Theorem II.** Consider the pair of Young functions

$$\tau(t) = e^t - t - 1, \quad \tau^*(s) = \begin{cases} (s+1) \log(s+1) - s & \text{if } s \geq -1, \\ +\infty & \text{otherwise,} \end{cases}$$

which are dual to each other. If  $\Lambda$  is a finite positive measure on some (locally) compact space  $S$ , we define for any measurable function  $h$  on  $S$  its Luxemburg norm by

$$\|h\|_\theta = \inf \left\{ a > 0 \mid \int \theta\left(\frac{|h|}{a}\right) d\Lambda \leq 1 \right\},$$

where  $\theta$  equals  $\tau$  or  $\tau^*$ . Correspondingly we obtain two Orlicz spaces  $(L^\tau, \|\cdot\|_\tau)$  and  $(L^{\tau^*}, \|\cdot\|_{\tau^*})$ , which are Banach spaces in topological duality; see e.g. Rao and Ren [RR]. Let  $\mathcal{C}$  denote a linear subspace of  $C_b(S)$ , and define the convex functional

$$\Gamma(f) := \int \tau(f) d\Lambda \quad (f \in \mathcal{C}).$$

If  $\ell$  is an element of the algebraic dual  $\mathcal{C}^\#$  of  $\mathcal{C}$  such that  $\Gamma^*(\ell) := \sup_{f \in \mathcal{C}} [\ell(f) - \Gamma(f)] < \infty$ , then there is some function  $K \in L^{\tau^*}$  such that

$$\ell(f) = \int fK d\Lambda \quad \forall f \in \mathcal{C} \quad \text{and} \quad \Gamma^*(\ell) = \int \tau^*(K) d\Lambda. \quad (3.1)$$

In particular  $K \geq -1$   $\Lambda$ -a.e. on  $S$ . This has been proved in Sections 5, 6 and 7 of Léonard [L1]. See in particular the proof of Theorem 7.1 of [L1]. There (3.1) has been stated in the case where the measure  $\Lambda$  has a certain form and where  $S = [0, T] \times (\mathbb{R}^d)^2 \times ((\mathbb{R}^d)^2 \setminus \{(0, 0)\})$ . However, inspecting the proof one finds that actually only the facts that  $S$  is a locally compact metric space and that  $\Lambda$  is a positive finite measure enter the proof.

To apply the result (3.1) in our situation, suppose first that  $\mathbb{E}$  is compact, and fix a path  $\eta$  such that  $S_\nu(\eta) < \infty$ . Then define the compact metrizable space  $S := I \times \mathcal{M} \times \mathbb{E}^2$ , the linear space  $\mathcal{C} := \{g \in C(S) \mid g(t, \mu, x, y) = Df(t, \mu, x, y) \text{ for some } f \in \mathcal{F}\}$ , and the positive finite measure  $\Lambda(dt, d\mu, dx, dy) := dt \bar{m}(\eta_t, d\mu, dx, dy)$ . Since

$$\mathcal{H}(\eta_s, f) = \ll \eta_s, \mathcal{G}_m(\eta_s) f_s \gg + \int d\bar{m}(\eta_s) \tau(Df_s), \quad (3.2)$$

we have that

$$J_T(\eta, f) = \tilde{\ell}(f) - \int_0^T ds \int d\bar{m}(\eta_s) \tau(Df_s) = \tilde{\ell}(f) - \Gamma(Df),$$

with  $\tilde{\ell}(f)$  denoting the linear functional

$$\tilde{\ell}(f) := \ll \eta_T, f_T \gg - \ll \eta_0, f_0 \gg - \int_0^T \ll \eta_s, \partial_s f_s + \mathcal{G}_m(\eta_s) f_s \gg ds. \quad (3.3)$$

Since

$$\begin{aligned} \frac{1}{\|Df\|_\tau} \tilde{\ell}(f) &= \tilde{\ell}\left(\frac{f}{\|Df\|_\tau}\right) \\ &= J_T(\eta, f) + \int_0^T ds \int d\bar{m}(\eta_s) \tau\left(\frac{Df_s}{\|Df_s\|_\tau}\right) \\ &\leq S_\nu(\eta) + 1 < \infty, \end{aligned}$$

we have  $\tilde{\ell}(f) = 0$  whenever  $Df = 0$  and hence  $\tilde{\ell}$  only depends on  $Df$ , i.e.  $\tilde{\ell}(f) = \ell(Df)$ , for some linear functional  $\ell$  on  $\mathcal{C}$ . With the above notation,  $S_\nu(\eta) = \Gamma^*(\ell)$ . Hence (3.1) gives us

$$S_\nu(\eta) = \int \tau^*(K) d\Lambda \quad \text{and} \quad \tilde{\ell}(f) = \int Df \cdot K d\Lambda \quad \forall f \in \mathcal{F},$$

for some function  $K \in L^{\tau^*}$ . Hence with  $g := K + 1 \geq 0$  we get the representation (2.5) and  $\eta \in H_\nu$ , where (2.3) follows from (3.3).

If  $\mathbb{E}$  is not compact, let  $\overline{\mathbb{E}}$  denote its completion with respect to the metric  $d$  chosen in Section 1. Then  $\overline{\mathbb{E}}$  is compact and  $C_u(\mathbb{E})$  is isomorphic to  $C(\overline{\mathbb{E}})$ . Thus every  $f \in \mathcal{F}_0$  can be regarded as a function on  $\mathcal{M}(\overline{\mathbb{E}})$ , and  $\mathcal{M}$  is naturally embedded into  $\mathcal{M}(\overline{\mathbb{E}})$ . Define  $\overline{S} = I \times \mathcal{M}(\overline{\mathbb{E}}) \times \overline{\mathbb{E}}^2$  and  $\Lambda$  as above. Then we get a function  $g$  defined on the bigger space  $\overline{S}$ , but  $\Lambda$  only charges the set  $S = I \times \mathcal{M} \times \mathbb{E}^2$ , i.e. we can set  $g \equiv 0$  on  $\overline{S} \setminus S$ , and conclude as above.

Now suppose on the other hand that  $\eta \in H_\nu$ . Then, since  $\eta$  solves the weak McKean-Vlasov equation with kernel  $dk_s = g_s dm$ , we conclude with Young's inequality that

$$\begin{aligned} J_T(\eta, f) &= \int_0^T \ll \eta_s, \mathcal{G}_{k_s}(\eta_s) f_s - \mathcal{G}_m(\eta_s) f_s \gg - \int \tau(Df_s) d\overline{m}(\eta_s) ds \\ &= \int_0^T \int (g_s - 1) Df_s - \tau(Df_s) d\overline{m}(\eta_s) ds \\ &\leq \int_0^T ds \int d\overline{m}(\eta_s) \tau^*(g_s - 1), \end{aligned}$$

which is finite by assumption. Therefore  $S_\nu(\eta) < \infty$  and Theorem II is proved.  $\square$

**Proof of Theorem III.** First we will prove the differentiability assertion. Using the elementary inequality

$$ab \leq e^a + b \log b - b \quad (a, b \geq 0)$$

we find that, for every  $f \in \mathcal{F}_0$  and  $\eta \in H_\nu$ ,

$$\begin{aligned} | \ll \eta_t, f \gg - \ll \eta_s, f \gg | &\leq \int_s^t dr \int d\overline{m}(\eta_r) |Df_r \cdot g_r| \\ &\leq (t - s) \|\gamma\| \cdot e^{\|f\|_{\mathcal{F}_0}} + \int_s^t dr \int d\overline{m}(\eta_r) (g_r \log g_r - g_r), \end{aligned} \quad (3.4)$$

where  $\|\gamma\|$  denotes the sup-norm of the function  $\gamma$ . Taking the supremum over all  $f \in \mathcal{F}_0$  with  $\|f\|_{\mathcal{F}_0} \leq 1$  we get that, for almost every  $t \in I$ ,

$$\limsup_{h \rightarrow 0} \left\| \frac{1}{h} (\eta_{t+h} - \eta_t) \right\|_{\mathcal{F}'_0} \leq \|\gamma\| e + \int d\overline{m}(\eta_t) (g_t \log g_t - g_t) < \infty.$$

Thus  $\frac{1}{h} (\eta_{t+h} - \eta_t)$  is weakly compact in  $\mathcal{F}'_0$  by the Alaoglu-Bourbaki theorem, for each  $t$  outside a Lebesgue null set  $N_0$ .

Now we prove that, for almost every  $t \in I$ , this net can have at most one accumulation point. Indeed, for each  $f \in \mathcal{F}_0$ , there exists a Lebesgue null set  $N(f) \subset I$  such that  $t \mapsto \ll \eta_t, f \gg$  is differentiable outside  $N(f)$ . If we already knew that there exists a countable dense set  $D \subset \mathcal{F}_0$ , then the limit of  $\frac{1}{h}(\eta_{t+h} - \eta_t)$  ( $h \rightarrow 0$ ) would be uniquely determined for every  $t$  outside  $N_0 \cup \bigcup_{f \in D} N(f)$ .

To prove separability of  $\mathcal{F}_0$ , let  $F_k$  denote a countable dense set of  $C_c^1(\mathbb{R}^k)$  with respect to the norm  $\|\varphi\| + \|\nabla\varphi\|$  ( $\varphi \in C_c^1(\mathbb{R}^k)$ ,  $k = 1, 2, \dots$ ), and choose  $G \subset C_u(\mathbb{E})$  denoting a countable dense set with respect to the sup-norm. Such a set  $G$  exists by Lemma II.6.3 of Parthasarathy [Pa]. Then the set  $D$  of functions  $f$  of the form (1.7) with  $\varphi \in F_k$ ,  $\psi_1, \dots, \psi_k \in G$  and  $k \in \mathbb{N}$  is dense in  $\mathcal{F}_0$ .

The above arguments imply that  $t \mapsto \eta_t \in \mathcal{F}'_0$  is almost everywhere weakly differentiable with weak derivative  $\dot{\eta}_t$  ( $t \in I$ ). Moreover, (3.4) shows that  $t \mapsto \eta_t \in \mathcal{F}'_0$  is weakly absolutely continuous and of strong bounded variation. Thus we infer from Theorem 3.8.6 in Hille and Phillips [HP] that  $\eta$  is the indefinite Bochner integral of  $\dot{\eta}$ :

$$\eta_t = \eta_0 + \int_0^t \dot{\eta}_s ds \quad (t \in I).$$

Corollary 2 of Theorem 3.8.5 in [HP] now states that  $\eta$  is even strongly differentiable.

Now let us turn to the proof of the Lagrangian representation formula of  $S_\nu$ . It is based on the following integration by parts formula: For each  $t \geq 0$  and  $f \in \mathcal{F}$ ,

$$\ll \eta_t, f_t \gg - \ll \eta_0, f_0 \gg = \int_0^t \ll \dot{\eta}_s, f_s \gg ds + \int_0^t \ll \eta_s, \partial_s f_s \gg ds. \quad (3.5)$$

This formula can be proved as Lemma 4.3 of [DG2], using the fact that  $t \mapsto \eta_t \in \mathcal{F}'_0$  is of strong bounded variation. Now, it follows from the definition of  $\mathcal{L}$  and Eq. (3.5) that, for any  $f \in \mathcal{F}$ ,

$$\int_0^T \mathcal{L}_t(\dot{\eta}_t, \eta_t) dt \geq \int_0^T (\ll \dot{\eta}_t, f_t \gg - \mathcal{H}(\eta_t, f_t)) dt = J_T(\eta, f).$$

Hence,

$$\int_0^T \mathcal{L}_t(\dot{\eta}_t, \eta_t) dt \geq S_\nu(\eta).$$

Next, it follows from Eq. (3.2) and the weak McKean–Vlasov equation that, for  $f \in \mathcal{F}_0$  and almost all  $t \in I$ ,

$$\begin{aligned} \ll \dot{\eta}_t, f \gg - \mathcal{H}(\eta_t, f) &= \int (g_t - 1)Df - \tau(Df) d\bar{m}(\eta_t) \\ &\leq \int \tau^*(g_t - 1) d\bar{m}(\eta_t), \end{aligned}$$

where we again used Young's inequality. Hence, by (2.5),

$$\int_0^T \mathcal{L}_t(\dot{\eta}_t, \eta_t) dt \leq S_\nu(\eta),$$

which finishes the proof of Theorem III.  $\square$

## 4. A Large Deviation Upper Bound.

**4.1. An exponential martingale.** For  $\eta \in D(I, \mathcal{M}^{II})$  and  $f \in \mathcal{F}$ , let  $J^{M,N}(\eta, f)$  denote the following functional

$$\begin{aligned} J_t^{M,N}(\eta, f) = & \ll \eta_t, f_t \gg - \ll \eta_0, f_0 \gg - \int_0^t \ll \eta_s, \partial_s f_s \gg ds \\ & - \frac{1}{MN} \int_0^t \mathcal{H}^{M,N}(\eta_s, MN f_s) ds. \end{aligned} \quad (4.1)$$

In view of Eqs. (1.8), (1.9) and (2.2), it is easily checked that, for each  $f \in \mathcal{F}$  and  $t \in I$ ,

$$\sup_{\eta \in D(I, \mathcal{M}^{II})} \left| J_t^{M,N}(\eta, f) - J_t(\eta, f) \right| \longrightarrow 0 \quad (M, N \rightarrow \infty). \quad (4.2)$$

From now on let us denote the coordinate process on  $D(I, \mathcal{M}^{II})$  with  $(\xi_t)_{t \geq 0}$ .

**Lemma 4.2.** *For each  $f \in \mathcal{F}$ , the process*

$$Z_t^{M,N,f} = \exp\{MN J_t^{M,N}(\xi, f)\}, \quad t \in I,$$

*is a  $(\mathcal{P}^{M,N}, \mathcal{F}_t)$ -martingale.*

*Proof.* Let  $g$  denote the function  $MNf$  for a given  $f \in \mathcal{F}$ . Then the process

$$M_t := e^{\ll \xi_t, g_t \gg} - \int_0^t e^{\ll \xi_s, g_s \gg} \mathcal{H}^{M,N}(\xi_s, g_s) ds \quad (t \in I)$$

is a martingale by definition of  $\mathcal{H}^{M,N}$  and by the boundedness of  $g$ . Integration by parts now yields

$$dZ_t^{M,N,f} = \exp \left[ - \ll \xi_0, g_0 \gg - \int_0^t \ll \xi_s, \partial_s g_s \gg + \mathcal{H}^{M,N}(\xi_s, g_s) ds \right] dM_t.$$

Therefore  $Z^{M,N,f}$  is a local martingale. But, again by boundedness,  $Z^{M,N,f}$  is even a true martingale.  $\square$

## 4.2. Exponential tightness.

The purpose of this section is to prove the exponential tightness property. It is stated in the following

**Proposition 4.4.** *For every  $L > 0$ , there exists a compact set  $K_L$  of  $D(I, \mathcal{M}^{II})$  such that*

$$\limsup_{M, N \rightarrow \infty} \frac{1}{MN} \log \mathcal{P}^{M,N}[\xi \notin K_L] \leq -L. \quad (4.3)$$

To prove the proposition, we will apply Theorem A in the Appendix with

$$\mathbb{F} = \{ \ll \cdot, f \gg \mid f \in \mathcal{F}_0 \}.$$

Clearly,  $\mathbb{F}$  is additive, separates points, and is contained in  $C_b(\mathcal{M}^{II})$ . The first condition of Theorem A will be shown in Section 4.2.2 and the second one will be proved in Section 4.2.1 below.

### 4.2.1. Exponential Tightness of $\ll \xi(\cdot), f \gg$ .

We will show that, for a fixed  $f \in \mathcal{F}_0$ , the distributions of the process  $\ll \xi(\cdot), f \gg$  under  $\mathcal{P}^{M,N}$ ,  $M, N \in \mathbb{N}$  are exponentially tight in  $D(I, \mathbb{R})$ . To this end, we will apply the following exponential version of Aldous' criterion given in Theorem B of Pukhalskii [Pu]:

**Theorem (Pukhalskii).** *Let  $(X^n)_{n \in \mathbb{N}}$  denote a sequence of stochastic processes with sample paths in  $D(I, \mathbb{R})$ , and suppose that*

- (i)  $\lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P \left[ \sup_{t \in I} |X_t^n| > L \right] = -\infty$
- (ii)  $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\tau} \frac{1}{n} \log P \left[ \sup_{t \leq \delta, t+\tau \in I} |X_{t+\tau}^n - X_\tau^n| > \epsilon \right] = -\infty \quad \forall \epsilon > 0,$

where the supremum in (ii) is taken over all stopping times  $\tau \leq T$ . Then the sequence  $(X^n)$  is exponentially tight in  $D(I, \mathbb{R})$ .

Its condition (i) follows from the fact that  $0 \leq \ll \xi, f \gg \leq \|f\|$ , where  $\|g\|$  denotes the sup-norm of a function  $g$ , regardless of the space on which  $g$  is defined.

Now let us prove condition (ii). To this end, let  $\mathcal{P}_\zeta^{M,N}$  denote the solution of the martingale problem associated with (1.4), where the process  $\xi$  starts from some  $\zeta \in \mathcal{M}^{II}$  that has the form (1.2). Then, by Assumption II,

$$\mathcal{H}^{M,N}(\xi_s, MN\beta f) \leq \|\gamma\| MN \exp\left(\|Df\|\beta\right),$$

for all  $\beta > 0$ . Choosing  $\beta = -\log \delta / \|Df\|$  and applying Doob's inequality, we get, with  $Z_t^{M,N,f}$  denoting the martingale of Lemma 4.2, that

$$\begin{aligned} \mathcal{P}_\zeta^{M,N} \left[ \sup_{t \leq \delta} (\ll \xi_t, f \gg - \ll \xi_0, f \gg) \geq \epsilon \right] &\leq \\ &\leq \mathcal{P}_\zeta^{M,N} \left[ \sup_{t \leq \delta} Z_t^{M,N,\beta f} \geq \exp\left(\epsilon MN\beta - \int_0^\delta \mathcal{H}^{M,N}(\xi_s, MN\beta f) ds\right) \right] \\ &\leq \exp\left(MN(\|\gamma\| + \frac{\epsilon}{\|Df\|} \log \delta)\right). \end{aligned}$$

Let us for the moment denote by  $\mathcal{M}_{M,N}^{II}$  the set of all two-level empirical measures of the form (1.2), for given  $M$  and  $N$ . Then, if  $\tau \leq T$  is a fixed stopping time, the strong Markov property yields that

$$\begin{aligned} \frac{1}{MN} \log \mathcal{P}^{M,N} \left[ \sup_{t \leq \delta, t+\tau \in I} (\ll \xi_{\tau+t}, f \gg - \ll \xi_\tau, f \gg) \geq \epsilon \right] &\leq \\ &\leq \frac{1}{MN} \log \sup_{\zeta \in \mathcal{M}_{M,N}^{II}} \mathcal{P}_\zeta^{M,N} \left[ \sup_{t \leq \delta} (\ll \xi_t, f \gg - \ll \xi_0, f \gg) \geq \epsilon \right] \\ &\leq \|\gamma\| + \frac{\epsilon}{\|Df\|} \log \delta, \end{aligned}$$

which tends to  $-\infty$  as  $\delta \rightarrow 0$ . Condition (ii) of the above Theorem now follows by applying

$$\mathcal{P}^{M,N} \left[ \sup_{t \leq \delta, t+\tau \in I} |Y_t^f| \geq \epsilon \right] \leq \mathcal{P}^{M,N} \left[ \sup_{t \leq \delta, t+\tau \in I} Y_t^f \geq \epsilon \right] + \mathcal{P}^{M,N} \left[ \sup_{t \leq \delta, t+\tau \in I} Y_t^{-f} \geq \epsilon \right]$$

with  $Y_t^f := \ll \xi_{\tau+t}, f \gg - \ll \xi_\tau, f \gg$  ( $f \in \mathcal{F}_0$ ).  $\square$

### 4.2.2. The compact containment condition.

We have to prove that, for every  $L > 0$ , there exists a compact set  $A_L \in \mathcal{M}^{II}$  such that

$$\limsup_{M,N \rightarrow \infty} \frac{1}{MN} \log \mathcal{P}^{M,N} [\exists t \leq T; \quad \xi_t \notin A_L] \leq -L. \quad (4.4)$$

Observe first that a set  $A \subset \mathcal{M}^{II}$  is compact if and only if the set

$$C := \left\{ \int \mu \zeta(d\mu) \mid \zeta \in A \right\}$$

is compact in  $\mathcal{M}$ . Indeed, if  $C$  is compact, Prohorov's theorem implies the existence of compact sets  $K_k \subset \mathbb{E}$  such that

$$\sup_{\mu \in C} \mu(K_k^c) \leq \epsilon 2^{-k}/k \quad (k = 1, 2, \dots),$$

whenever  $\epsilon > 0$  is given. Define  $F_\epsilon = \bigcap_{k=1}^{\infty} \left\{ \mu \in \mathcal{M} \mid \mu(K_k^c) \leq 1/k \right\}$ . Then  $F_\epsilon$  is compact, again by Prohorov's theorem, and

$$\zeta(F_\epsilon^c) \leq \sum_{k=1}^{\infty} \zeta(\{\mu \mid \mu(K_k^c) > 1/k\}) \leq \sum_{k=1}^{\infty} k \int \mu(K_k^c) \zeta(d\mu),$$

which is less than  $\epsilon$ , for each  $\zeta \in A$ . Thus  $A$  is tight and hence compact. The converse statement follows from the continuity of the map  $\zeta \mapsto \int \mu \zeta(d\mu)$  ( $\zeta \in \mathcal{M}^{II}$ ).

Therefore to prove (4.4) it suffices for us to show that, for every  $L > 0$ , there exists a compact  $C_L \subset \mathcal{M}$  such that

$$\limsup_{M,N \rightarrow \infty} \frac{1}{MN} \log \mathcal{P}^{M,N} \left[ \exists t \leq T \text{ with } \int \mu \xi_t(d\mu) \notin C_L \right] \leq -L. \quad (4.5)$$

To this end, fix  $M$  and  $N$  and define, for  $\mathbf{x} = (x^{ij}) \in \mathbb{E}^{MN}$ , the functions  $\bar{\gamma}^{ij}(\mathbf{x})$  ( $i = 1, \dots, M, j = 1, \dots, N$ ) by (1.3). Consider the generator  $\bar{A}^{M,N}$  defined for bounded and measurable functions  $F$  on  $\mathbb{E}^{MN}$  by

$$\bar{A}^{M,N} F(\mathbf{x}) = \sum_{\substack{i=1, \dots, M \\ j=1, \dots, N}} \int \left( F_{ij}(\mathbf{x}, y) - F(\mathbf{x}) \right) \bar{\gamma}^{ij}(\mathbf{x}) \pi(x^{ij}, dy), \quad (4.6)$$

where  $F_{ij}(\mathbf{x}, y)$  denotes  $F$  evaluated at that point of  $\mathbb{E}^{MN}$  where the component  $x^{ij}$  of  $\mathbf{x}$  is replaced by  $y$ . As in Section 1.1 one sees that the martingale problem associated with  $\bar{A}^{M,N}$  is well posed, and that the unique solution  $\mathbf{X}_t = (X_t^{ij})$  ( $t \geq 0$ ) is a càdlàg strong Markov process. By projecting  $\mathbf{X}$  via (1.1) and (1.2) onto  $\mathcal{M}^{II}$  we recover the solution of the martingale problem  $\xi^{M,N}$  associated with (1.4). On the other hand, modulo permutations of the indices,  $\mathbf{X}$  can be recovered from  $\xi^{M,N}$ . Therefore we can assume that  $\mathcal{P}^{M,N}$  arises as the law of the projection of a stochastic process  $\mathbf{X}_t$  ( $t \geq 0$ ), that is defined on a probability space  $(\Omega, \mathcal{B}, \bar{P})$ , and that solves the martingale problem for  $\bar{A}^{M,N}$ .

For a compact set  $K \subset \mathbb{E}$ , let  $\tau_K^{ij}$  denote the stopping times

$$\tau_K^{ij} = \inf \{ t \geq 0 \mid X_t^{ij} \notin K \}. \quad (4.7)$$

We will prove below

**Lemma 4.5.** *For every  $R > 0$ , there exists a compact set  $K_R$  in  $\mathbb{E}$ , independent of  $M$  and  $N$ , and a family  $(\sigma_{K_R}^{ij})$  of stopping times ( $1 \leq i \leq M$ ,  $1 \leq j \leq N$ ) such that*

- (i)  $\sigma_{K_R}^{ij} \leq \tau_{K_R}^{ij}$  a.s.
- (ii) *The stopping times  $\sigma_{K_R}^{ij}$  are independent.*
- (iii)  $\bar{P}[\sigma_{K_R}^{ij} \leq T] \leq \exp(-R)$ .

Now put

$$C_L = \left\{ \mu \in \mathcal{M}; \quad \mu(K_{k^2L}^c) \leq \frac{1}{k} \quad \forall k \right\}.$$

Then  $C_L$  is compact in  $\mathcal{M}$ . Moreover, with  $\bar{\mu}(t) = \int \mu \xi_t(d\mu)$ ,  $t \in I$ , Lemma 4.5 yields

$$\begin{aligned} \mathcal{P}^{M,N} \left[ \exists t \leq T; \quad \bar{\mu}(t) \notin C_L \right] &\leq \sum_{k=1}^{\infty} \mathcal{P}^{M,N} \left[ \sup_{t \leq T} \bar{\mu}(t)(K_{k^2L}^c) \geq 1/k \right] \\ &\leq \sum_{k=1}^{\infty} \bar{P} \left[ \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M I_{\{\sigma_{K_{k^2L}}^{ij} \leq T\}} \geq 1/k \right] \\ &\leq \sum_{k=1}^{\infty} e^{-kLMN} \prod_{ij} \bar{E} \left[ \exp(kLI_{\{\tau_{K_{k^2L}}^{ij} \leq T\}}) \right] \\ &\leq 2^{MN} \frac{e^{-LMN}}{1 - e^{-LMN}}, \end{aligned}$$

which in turn implies (4.4) and hence (4.3).  $\square$

*Proof of Lemma 4.5.* Define

$$r^{ij}(t) = \inf \left\{ s \geq 0; \quad \frac{1}{\|\gamma\|} \int_0^s \gamma^{ij}(\mathbf{X}_u) du \geq t \right\}$$

and let  $\tilde{X}_t^{ij} = X^{ij}(r^{ij}(t))$ . It follows that the process  $\tilde{\mathbf{X}}_t = (\tilde{X}_t^{ij})$  is a solution of the martingale problem for an operator having the form (4.6), but with the function  $\gamma^{ij}$  replaced by the constant  $\|\gamma\|$ . Uniqueness of this martingale problem now implies that the processes  $\tilde{X}^{ij}$  ( $i = 1, \dots, M$ ,  $j = 1, \dots, N$ ) are independent. Now, we can define

$$\sigma_K^{ij} = \inf \{ t \geq 0 \mid \tilde{X}_t^{ij} \notin K \}.$$

Since  $\gamma \leq \|\gamma\|$ , it follows that  $r^{ij}(t) \geq t$ , and hence  $\sigma_{K_R}^{ij} \leq \tau_{K_R}^{ij}$ . This proves assertions (i) and (ii).

To see that we can choose  $K_R$  such that (iii) holds, observe first that we can find an integer  $q$  such that the number of jumps of the process  $\tilde{X}^{ij}$  until time  $T$  does not exceed  $q$  with probability  $1 - e^{-2R}$ . By Prohorov's Theorem, we can now choose a compact  $K_R^1$  so that a Markov chain starting from some  $x_0 \in \mathbb{E}$  and moving according to the transition probability  $\pi$  does not leave  $K_R^1$  in less than  $q$  steps with probability  $1 - e^{-2R}$ . By the Feller property of the kernel  $\pi$ , we can even find, for each compact  $K_0 \subset \mathbb{E}$ , a compact  $K_R$  such that the above holds uniformly for each  $x_0 \in K_0$ . Since our starting points  $\nu^{M,N}$  are assumed to converge weakly, (iii) now follows easily by an other application of Prohorov's Theorem.  $\square$

### 4.3 Upper bound.

For  $\delta > 0$  and  $\eta \in D(I, \mathcal{M}^{II})$  define

$$S_\nu^\delta(\eta) = \begin{cases} S_\nu(\eta) - \delta, & \text{if } S_\nu(\eta) < \infty, \\ 1/\delta, & \text{otherwise.} \end{cases}$$

By standard arguments and Proposition 4.4 (see e.g. Dembo and Zeitouni [DZ], p. 6), the upper bound follows from the following

**Proposition 4.6.** *For every  $\delta > 0$  and each  $\eta^0 \in D(I, \mathcal{M}^{II})$  there exists an open neighbourhood  $V$  of  $\eta^0$  such that*

$$\limsup_{M, N \rightarrow \infty} \frac{1}{MN} \log \mathcal{P}^{M, N}(V) \leq -S_\nu^\delta(\eta^0). \quad (4.8)$$

*Proof.* If  $\eta_0^0 \neq \nu$ , this is trivial. Therefore, we will assume  $\eta_0^0 = \nu$  in the sequel. By the definition of the rate function, there exists  $f_\delta \in \mathcal{F}$  such that

$$J_T(\eta^0, f_\delta) \geq S_\nu^{\delta/2}(\eta^0).$$

Let  $V$  denote the set

$$V = \left\{ \eta \in D(I, \mathcal{M}^{II}) \mid |J_T(\eta, f_\delta) - J_T(\eta^0, f_\delta)| < \delta/2 \right\}.$$

Since the mapping  $\eta \mapsto J_T(\eta, f_\delta)$  is continuous on  $D(I, \mathcal{M}^{II})$ ,  $V$  is an open neighborhood of  $\eta^0$ . Lemma 4.2 gives us that

$$1 = E^{M, N} \left[ \exp \left( MN J_T^{M, N}(\xi, f_\delta) \right) \right].$$

Therefore,

$$-J_T(\eta^0, f_\delta) \geq \frac{1}{MN} \log E^{M, N} \left[ \exp \left( MN (J_T^{M, N}(\xi, f_\delta) - J_T(\eta^0, f_\delta)); \xi \in V \right) \right].$$

Thus, by adding and subtracting  $J_T(\xi, f_\delta)$  inside the exponential term and using the definition of  $V$ , it follows that

$$-J_T(\eta^0, f_\delta) \geq \frac{1}{MN} \log \mathcal{P}^{M, N}(V) - \delta/2 - \sup_{\zeta \in D(I, \mathcal{M}^{II})} \left| J_T^{M, N}(\zeta, f_\delta) - J_T(\zeta, f_\delta) \right|.$$

But,

$$\sup_{\zeta \in D(I, \mathcal{M}^{II})} \left| J_T^{M, N}(\zeta, f_\delta) - J_T(\zeta, f_\delta) \right| \longrightarrow 0 \quad (M, N \rightarrow \infty),$$

and (4.8) is proved.

### Appendix: A criterion for exponential tightness in $D(I, E)$

The purpose of this appendix is to give a criterion for the exponential tightness of a sequence  $X^1, X^2, \dots$  of stochastic processes taking values in some completely

regular topological space  $E$ . We shall assume that for each natural number  $n \in \mathbb{N}$  the process  $X^n$  induces a measurable mapping from some given probability space  $(\Omega, \mathcal{B}, P)$  into the Skorohod space  $D(I, E)$ , endowed with the Skorohod topology and its Borel field (cf. [Ja]). Note that this mapping is always measurable, if the process  $X^n$  is right continuous with left limits and the space  $E$  is at least separable and metrizable. Let  $(a_n)$  denote some sequence of positive real numbers increasing to infinity. We will say that the sequence  $(X^n)$  is *exponentially tight* in  $D(I, E)$  with speed  $(a_n)$ , if for each  $L > 0$  we can find a compact  $K \subset D(I, E)$  such that

$$\limsup_{n \uparrow \infty} \frac{1}{a_n} \log P[X^n \notin K] \leq -L.$$

The following criterion is an exponential version of a result due to Jakubowski; see [Ja], Theorem 3.1.

**Theorem A:** *Let  $E$  be a completely regular topological space with metrizable compacts. Then the sequence  $(X^n)$  is exponentially tight in  $D(I, E)$  with speed  $(a_n)$ , if and only if the following two conditions are fulfilled:*

(i) *For every  $M > 0$ , there exists a compact  $A_M \subset E$ , such that*

$$\limsup_{n \uparrow \infty} \frac{1}{a_n} \log P[\exists t : X_t^n \notin A_M] \leq -M.$$

(ii) *There exists an additive family  $\mathbb{F} \subset C(E)$ , separating the points of  $E$ , such that, for each  $f \in \mathbb{F}$ , the sequence  $(f(X^n))$  is exponentially tight in  $D(I, \mathbb{R})$  with speed  $(a_n)$ .*

**Remarks:**

1. The space  $C(I, E)$  of continuous paths, endowed with the usual compact open topology, is a closed topological subspace of  $D(I, E)$  (cf.[Ja], Proposition 1.6 (i)). This shows that the above criterion holds in  $C(I, E)$  as well.
2. The assumption  $\mathbb{F} \subset C(E)$  may be weakened. It suffices to assume that the restriction of each  $f \in \mathbb{F}$  on each set  $A_M$  is continuous.
3. The event  $\{\exists t \in I \mid X_t \notin A\}$  is Borel measurable if  $A \subset E$  is closed (cf.[Ja], Proposition 1.6 (v)).

**Proof of Theorem 1:** The necessity of the two conditions can be shown as in the case of ordinary tightness. To prove sufficiency, we note first that, by Lemma 3.2 in [Ja], the family  $\mathbb{F}$  may be assumed to be countable:  $\mathbb{F} = \{f_1, f_2, \dots\}$ . Now, for each  $k \in \mathbb{N}$  and every  $R > 0$ , there exists, by assumption, a compact  $\tilde{C}_R^k \subset D(I, \mathbb{R})$  with

$$\limsup_{n \uparrow \infty} \frac{1}{a_n} \log P[f_k(X^n) \notin \tilde{C}_R^k] \leq -R - 1.$$

Hence there is some  $n_0 \in \mathbb{N}$  such that, for all  $n \geq n_0$ ,

$$P[f_k(X^n) \notin \tilde{C}_R^k] \leq e^{-R a_n}.$$

Since  $D(I, \mathbb{R})$  is Polish, the set  $\tilde{C}_R^k$  may be enlarged to a compact  $C_R^k$ , satisfying

$$P\left[f_k(X^n) \notin C_R^k\right] \leq e^{-Ra_n} \quad \forall n \in \mathbb{N}.$$

Now let  $L > 1$  be given. We define

$$K_L = \left\{ w \in D(I, E) \mid w(t) \in A_L \forall t, f_k(w(\cdot)) \in C_{kL}^k \forall k \in \mathbb{N} \right\}.$$

In [Ja] it is shown that  $K_L$  is a compact subset of  $D(I, E)$ . So, to establish exponential tightness, it suffices to observe that

$$\begin{aligned} \limsup_{n \uparrow \infty} \frac{1}{a_n} \log P\left[X^n \notin K_L\right] &\leq (-L) \vee \left( \limsup_{n \uparrow \infty} \frac{1}{a_n} \log \sum_{k=1}^{\infty} P\left[f^k(X^n) \notin C_{kL}^k\right] \right) \\ &\leq (-L) \vee \left( \limsup_{n \uparrow \infty} \frac{1}{a_n} \log \sum_{k=1}^{\infty} e^{-kLa_n} \right) \\ &\leq -L. \end{aligned}$$

**Acknowledgement.** We would like to thank the referee of the paper for several insightful comments.

## REFERENCES

- [Bi] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Inc., New York, 1968.
- [D1] D. A. Dawson, *Asymptotic analysis of multilevel stochastic systems*, Stochastic differential systems (ed. M. Métivier and E. Pardoux). Lecture Notes in Control and Information Sciences, vol. 69, 1984, pp. 76-89.
- [D2] D. A. Dawson, *Stochastic ensembles and hierarchies*, Stochastic Processes and their Applications (ed. K. Itô and T. Hida). Lecture Notes in Math., vol. 1203, 1985, pp. 20-37.
- [DG1] D. A. Dawson, J. Gärtner, *Multilevel large deviations and interacting diffusions*, Probab. Theory Relat. Fields **98** (1994), 432-487.
- [DG2] D. A. Dawson, J. Gärtner, *Large deviations from the McKean-Vlasov limit for weakly interacting diffusions*, Stochastics **20** (1987), 247-308.
- [DZ] A. Dembo, O. Zeitouni, *Large Deviations Techniques*, Jones and Bartlett Publishers, Boston, London, 1993.
- [DK] B. Djehiche, I. Kaj, *The rate function for some measure-valued jump processes*, Annals of Probability **23** (1995), 1414-1438.
- [EK] S. Ethier, T. Kurtz, *Markov Processes, Characterization and Convergence*, Wiley-Interscience, New York, 1986.
- [HP] E. Hille, R. Phillips, *Functional Analysis and Semi-groups*, American Mathematical Society, Providence, 1957.

- [Ja] A. Jakubowski, *On the Skorohod topology* (1986), Ann. Inst. Henri Poincaré. **B22**, 263-285.
- [L1] C. Léonard, *On the large deviations for particle systems associated with spatially homogeneous Boltzmann type equations*, Probab. Theory Relat. Fields **101** (1995), 1-44.
- [L2] C. Léonard, *Large deviations for long range interacting particle systems with jumps*, Ann. Inst. Henri Poincaré **B31** (1995), 289-323.
- [Pa] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.
- [Pu] A. Pukhalskii, *The method of stochastic exponentials for large deviations*, Stoch. Process. Applications **54** (1994), 45-70.
- [RR] M. M. Rao, Z. D. Ren, *Theory of Orlicz Spaces*, Marcel Dekker, Inc., New York, 1991.

Boualem Djehiche  
 Division of Mathematical Statistics  
 The Royal Institute of Technology  
 S-100 44 Stockholm, Sweden  
 boualem@math.kth.se

Alexander Schied  
 Institut für Mathematik  
 Humboldt-Universität  
 Unter den Linden 6  
 D-10099 Berlin, Germany  
 schied@mathematik.hu-berlin.de